Title

White matter and literacy: A dynamic system in flux

Authors

Roy, Ethan

Richie-Halford, Adam

Kruper, John

et al.

Publication Date

2024-02-01

DOI

10.1016/j.dcn.2024.101341

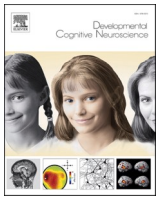
Peer reviewed

Developmental Cognitive Neuroscience 65 (2024) 101341

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: www.elsevier.com/locate/dcn



White matter and literacy: A dynamic system in flux

Ethan Roy [a,*], Adam Richie-Halford [a,b,c], John Kruper [b], Manjari Narayan [c], David Bloom [b], Pierre Nedelec [f], Andreas M. Rauschecker [f], Leo P. Sugrue [f], Timothy T. Brown [d], Terry L. Jernigan [e], Bruce D. McCandliss [a], Ariel Rokem [b], Jason D. Yeatman [a,c]

[a] Graduate School of Education, Stanford University, Stanford, CA, USA
[b] Department of Psychology and eScience Institute, University of Washington, Seattle, WA, USA
[c] Division of Developmental-Behavioral Pediatrics, Stanford University, Stanford, CA, USA
[d] School of Medicine, University of California San Diego, San Diego, CA, USA
[e] Center for Human Development, University of California San Diego, San Diego, CA, USA
[f] Department of Radiology and Biomedical Imaging, University of California San Francisco, San Francisco, CA, USA

ARTICLE INFO

Keywords:
White Matter
Reading
Longitudinal data
Diffusion MRI
Big Datasets

ABSTRACT

Cross-sectional studies have linked differences in white matter tissue properties to reading skills. However, past studies have reported a range of, sometimes conflicting, results. Some studies suggest that white matter properties act as individual-level traits predictive of reading skill, whereas others suggest that reading skill and white matter develop as a function of an individual's educational experience. In the present study, we tested two hypotheses: a) that diffusion properties of the white matter reflect stable brain characteristics that relate to stable individual differences in reading ability or b) that white matter is a dynamic system, linked with learning over time. To answer these questions, we examined the relationship between white matter and reading in a five-year longitudinal dataset and a series of large-scale, single-observation, cross-sectional datasets (N = 14,249 total participants). We find that gains in reading skill correspond to longitudinal changes in the white matter. However, in the cross-sectional datasets, we find no evidence for the hypothesis that individual differences in white matter predict reading skill. These findings highlight the link between dynamic processes in the white matter and learning.

1. Introduction

White matter - the tissue that contains the long-range axonal connections among different brain regions - was historically viewed as static infrastructure critical for healthy cognitive function (Wernicke, 1874; Geschwind, 1965). However, the field went through a dramatic shift as it became clear that oligodendrocytes, the glial cells responsible for myelination in the white matter, actively monitor neural activity (Pease-Raissi and Chan, 2021; Fields, 2015). Through signaling mechanisms that sense neuronal discharges, oligodendrocytes actively change properties of the white matter in response to fluctuations in neural activity (Ishibashi et al., 2006; Barres and Raff, 1993). Indeed, it is now largely appreciated that the white matter not only plays an essential role in behavior, but also that plasticity in the white matter is a critical component of the learning process (Fields, 2015).

Given that white matter is sculpted by experience, and that properties of the white matter can undergo dramatic changes over timescales ranging from hours (Gibson et al., 2014; Sagi et al., 2012), to days (Huber et al., 2021), to years (Yeatman et al., 2014; Lebel et al., 2008), we now must grapple with how to interpret individual differences in white matter structure. It is often assumed that white matter differences are static traits that explain differences in behavior and can serve as useful biomarkers of, for example, learning disabilities (Gabrieli, 2009; Langer et al., 2017; Vandermosten et al., 2015; Gabrieli et al., 2015). This perspective is supported by dozens of studies that have: a) observed correlations between white matter diffusion properties and behavioral measures of academic skills or b) found group differences in white matter diffusion properties between individuals with versus without learning disabilities (Ozernov-Palchik and Gaab, 2016; Vandermosten et al., 2012; Ben-Shachar et al., 2007).

Reading serves as a paradigmatic example of the relationship between white matter and academic abilities. Seminal research by

* Correspondence to: Stanford Graduate School of Education, 482 Galvez Mall, Stanford, CA 94305.
E-mail address: ethanroy@stanford.edu (E. Roy).

https://doi.org/10.1016/j.dcn.2024.101341
Received 22 February 2023; Received in revised form 24 August 2023; Accepted 3 January 2024
Available online 6 January 2024

Klingberg and colleagues (Klingberg et al., 2000) discovered differences in white matter diffusion properties between dyslexic and control participants, inspiring dozens of labs around the world to try to characterize the white matter phenotype of dyslexia (and reading abilities more broadly) (Ben-Shachar et al., 2007; Beaulieu et al., 2005; Niogi and McCandliss, 2006a; Deutsch et al., 2005; Saygin et al., 2013; Boets et al., 2013). Although much of this work has identified white matter differences between strong and struggling readers, these studies have reported a variety of, and sometimes even conflicting, results (see Supplemental Table 1 for an overview of past findings). These incongruencies also appear in various meta-analyses, with some identifying a link between reading ability and the white matter (Vandermosten et al., 2012) and others finding no brain-behavior relationship in the domain of reading (Moreau et al., 2018). Regardless of the direction of these effects, many of these studies have interpreted observed group differences in the white matter properties of dyslexic versus control participants as a static brain-behavior relationship that can be used to classify an individual based on intrinsic properties of their brain.

In contrast to studies exploring biomarkers that emphasize static brain-behavior relationships, longitudinal studies have highlighted the dynamic relationship between white matter plasticity and learning (Huber et al., 2021; Wandell and Yeatman, 2013; Yeatman et al., 2012a; Wang et al., 2017). These studies offer compelling evidence that rates of white matter development differ dramatically among individuals and are influenced by an individual's learning experiences. For example, Yeatman and colleagues (2012) showed that over a three year period participants with above-average reading skills showed increases in fractional anisotropy (FA), a quantity measured with diffusion MRI, in specific white matter pathways. Struggling readers, on the other hand, showed declines in FA in those same white matter pathways (Yeatman et al., 2012a). Similarly, Wang and colleagues (2017) showed divergent developmental trajectories for the core reading circuitry at the very beginning of reading instruction (Wang et al., 2017). Furthermore, intervention studies have identified dramatic changes in the diffusion properties following intensive reading interventions (Keller and Just, 2009; Huber et al., 2018). Even over the course of a couple weeks of intensive intervention, white matter tissue properties can change with large effect sizes, calling into question the stability of cross-sectional group comparisons (Huber et al., 2018). These studies challenge the assumption that white matter differences observed in poor readers are rooted in static brain traits, and instead suggest that the white matter is a dynamic, experience-dependent system that is linked with learning over development.

Until recently, the vast majority of published brain-behavior relationships have relied on small ($n < 50$), relatively homogenous participant samples, recruited from the surrounding community of a single research institution (Vandermosten et al., 2012). As highlighted by recent work (Marek et al., 2022), results from small and homogenous samples often do not generalize to a broader population. Furthermore, the relationship between white matter properties and reading skill has been shown to vary across socioeconomic groups (Gullick et al., 2016; Turesky et al., 2022; Ozernov-Palchik et al., 2019). Together, these findings not only suggest that demographic factors should be considered when exploring brain-behavior relationships but also raise questions as to how the link between socioeconomic status, reading, and white matter development unfolds at a population level. Resolving these issues has only recently become possible as the neuroimaging community has come together to produce large, publicly available, multi-site datasets that sample a much broader swath of the population than traditional, single-laboratory studies (Alexander et al., 2017; Jernigan et al., 2016; Casey et al., 2018).

For example, a recent cross-sectional analysis of the Healthy Brain Network (HBN) dataset (Alexander et al., 2017), representing the largest study of white matter and reading to date, did not reveal any significant differences in FA between struggling and typical readers nor a significant relationship between FA and reading scores (Meisler and Gabrieli, 2022). These findings not only raise questions about the observed relationships between reading and the white matter in smaller, cross-sectional studies, but also about the impact of sample makeup on these brain-behavior correlations, as the HBN dataset is a much more diverse sample than those found in traditional single-lab studies.

In the present study, we seek to test two hypotheses surrounding the relationship between white matter and academic skills: a) that individual differences in white matter diffusion properties reflect stable brain traits that relate to academic skills or b) that white matter and learning are dynamically linked over time and change in response to an individual's educational environment. To do so, we capitalize on a large-scale longitudinal dataset to explore the longitudinal relationship between changes in white matter properties and reading development. These data do not reveal any cross-sectional relationships between reading skill and white matter properties but, rather, show that gains in reading skill and changes in the white matter are linked longitudinally. This finding supports the hypothesis that white matter and learning are part of a dynamic system that changes as a function of educational experience.

We then leverage four additional large-scale public datasets, totaling more than 12,000 children and adults, to examine the cross-sectional correlations between white matter and reading ability. In this sample, which is more than two orders of magnitude larger than most previous studies, we find no support for the hypothesis that white matter features serve as static traits that predict differences in reading ability. Our analysis suggests that previous studies were detecting characteristics of small, demographically homogeneous samples. Together these findings suggest that a child's environment exerts a dramatic influence on white matter development and reject the notion that reading difficulties are explained by a stable white matter phenotype.

2. Materials and methods

To explore questions surrounding the properties of the white matter networks underlying reading skill, we leveraged five large-scale, publicly available datasets. These datasets included the Pediatric Longitudinal Imaging, Neurocognition, and Genetics (PLING (Wierenga et al., 2018)) dataset, the Child Mind Insitute's Healthy Brain Network (HBN) dataset (Alexander et al., 2017), the University of California San Diego's Pediatric, Imaging, Neurocognition, and Genetics (PING (Jernigan et al., 2016), and the Human Connectome Project's (HCP) Young Adult dataset (Harms et al., 2018). All five of these datasets provide neuroimaging data, demographic information, and a variety of behavioral assessments, including raw and age-standardized reading. We attempted to be as consistent as possible in our analytic pipelines, but due to differences in the available measures across these datasets, this was not always possible. An overview of the various demographic and behavioral measures included in each dataset are provided in Table 1.

Furthermore, due to differences in the processed data derivatives made available across the various datasets, the diffusion MRI data for the HBN, ABCD, and HCP-YA samples were analyzed using pyAFQ (Kruper et al., 2021) and the PLING and PING samples were analyzed using AtlasTrack (Hagler et al., 2009). Although tractometry was performed using different computational pipelines, past studies have shown that these analyses are robust to the details of the methodology (Kruper et al., 2021). Code to reproduce the results and figures is available at https://github.com/earoy/longitudinal_wm.

2.1. PLING

The PLING dataset was used to explore the relationship between changes in white matter properties and changes in academic skills over time. This dataset tracked 176 individuals over the course of 5 years, although not all participants completed all 5 time points. After excluding participants who did not participate in at least 3 time points and those that did not pass quality control, we were left with a sample of 73

Table 1
Overview of the neuroimaging, demographic, and reading measures available in each dataset.

Dataset	Number of Observations per Participant	Tractometry Pipeline	Demographic Measures	Reading Measure
PLING n = 73	3-5	AtlasTrack	n/a	Raw TOWRE
HBN n = 777	1	pyAFQ	Parental Income	Age-adjusted WIAT, Age-adjusted TOWRE
PING n = 1119	1	AtlasTrack	Parental Income	Age-adjusted NIH Toolbox
ABCD n = 11,080	1	pyAFQ	Parental Income, Neighborhood Deprivation, School Achievement	Age-adjusted NIH Toolbox
HCP-YA n = 1200	1	pyAFQ	n/a	Age-adjusted NIH Toolbox

individuals. At the time of the first observation, participants were between 5 and 10 years old.

At each time point, participants completed an MRI scanning session as well as an assortment of behavioral tasks. dMRI data was collected at each time point and diffusion MRI metrics from each timepoint were provided based on AtlasTrack. The available AtlasTrack metrics for this dataset include mean FA and MD values for 37 major white matter tracts. The behavioral measures we used in this analysis included a composite reading raw score derived from the Test of Word Reading Efficiency (TOWRE) Sight Word Efficiency and Phonetic Decoding Efficiency subtests (Torgesen et al., 2011). We choose raw TOWRE score, as using an age-normalized score will potentially obscure within-individual growth from year to year.

Based on past research investigating the link between reading and white matter (Wandell and Yeatman, 2013), we centered these analyses on the left arcuate fasciculus, inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), the right homologues of these tracts, and the frontal and posterior corpus callosum. However, because our initial longitudinal models only revealed a relationship between reading and FA in the left arcuate, we limited our subsequent modeling approaches to only the left arcuate.

To explore the longitudinal dynamics between reading skill and diffusion properties of the white matter, we constructed a series of longitudinal models. Although most participants (81%) participated in consecutive timepoints, some skipped an observation and therefore had missing data. We used t-tests to compare the reading scores and FA values of participants who dropped out of the study and those who did not. These tests suggested that missingness in the data was independent of both reading and white matter properties (both $p > 0.05$). Because these data were missing at random, we relied on Full Information Maximum Likelihood to generate estimations for these missing values while fitting the parallel-process latent growth model. For the linear-mixed effects and mlVAR models, missing observations were dropped from the analysis.

First, we generated a linear-mixed effects model predicting mean-centered TOWRE reading scores from time point, initial age, mean-centered FA at each time point, and overall mean FA as fixed effects, and participants as a random effect. The use of mean-centered FA allowed us to discern year-to-year change in FA from overall FA within each participant. To better understand the time series of FA and reading score change, we used a multi-level vector autoregression model (mlVAR (Epskamp et al., 2018a); Bringmann et al., 2013). This model used time series data to test if FA at a given time point predicted reading scores at the subsequent time point or vice versa. Finally, we generated a parallel-process latent growth curve model to understand the relationship between the rate of reading score change and the rate of FA development. We constructed this latent growth curve model using the lavaan package in R (Rosseel, 2012).

2.2. HBN

The HBN dataset consists of diffusion MRI and phenotypic data from over 1500 participants ranging from 5 to 21 years of age from the greater New York City area. In the HBN sample, neuroimaging data were collected at four different scanner sites, however the present analysis only included data from the three sites that used 3 T scanners: Rutgers University Brain Imaging Center, the CitiGroup Cornell Brain Imaging Center, and the CUNY Advanced Science Research Center. The sample analyzed in the present study consisted of 777 participants who passed quality control and had both neuroimaging data and various phenotypic measures, including reading scores and socioeconomic status, as indexed by parental income. To assess reading reading skill, we use the TOWRE age standardized total score. We also calculated an age standardized composite score based on the word reading and pseudoword decoding subsections of the Wechsler Individual Achievement Test (WIAT (McCrimmon and Climie, 2011)) as a second measure of reading skill. Individuals were labeled as struggling readers if their composite score was below 85 (one standard deviation from the mean) on either assessment. The use of age standardized scores allowed us to control for age-related gains in reading and to generate reading groups that spanned the age range of the entire sample.

The raw diffusion MRI data were preprocessed using qsiprep (Cieslak et al., 2020) and quality control was performed on the entire dataset using a neural network trained on ratings generated by a combination of diffusion imaging experts and community scientist on two subsets of the HBN dataset (Richie-Halford et al., 2022a). Any participants who did not meet quality control standards were excluded from the analysis. Additionally, qsiprep outputs measures of MRI data quality such as framewise-displacement and neighborhood voxel correlations. We entered these two data quality measures as covariates into our models, as individual differences in in-scanner motion, even among individuals who pass gross quality control, can have an impact on derived metrics such as fractional anisotropy (FA).

Once the diffusion imaging data were preprocessed, pyAFQ was used to calculate tractometry properties (Kruper et al., 2021). Briefly, constrained spherical deconvolution (Tournier et al., 2007), implemented in DIPY (Garyfallidis et al., 2014) was used to estimate fiber orientation distributions in every voxel, and probabilistic tractography was used to propagate streamlines throughout the white matter. 24 major white matter tracts were identified as originally described by Yeatman and colleagues (2012) (Yeatman et al., 2012b). Each tract was sampled to 100 nodes. Fractional anisotropy (FA), mean diffusivity (MD), and mean kurtosis (MK) were calculated at each node using the Diffusional Kurtosis Model (DKI (Jensen et al., 2005; Henriques et al., 2021) and axonal water fraction (AWF) was calculated using the White Matter Tract Integrity Model (WMTI (Fieremans et al., 2011)).

To assess the extent to which our computational pipeline impacted the results, we also implemented a pipeline that generated the streamlines using anatomically-constrained tractography (Smith et al., 2012) implemented in MRTRIX3 (Tournier et al., 2019). Tract identification

proceeded as above. We then grouped the data based on the behavioral data to make group comparisons across the various tracts. These diffusion features were also used as input to machine learning algorithms to predict reading scores. The results of both tractography pipelines were similar and, ultimately, the DIPY tractography was used.

With neuroimaging data, there are often differences between scanners, which can create variation in data quality and the scale of diffusion properties. This makes comparing images acquired from different sites challenging. To account for between-scanner variation, ComBat harmonization was performed on the tractometry data to correct for any scanner-related variance from the data (Fortin, 2017, 2018; Johnson et al., 2007). We also made sure to "protect" potential effects of age, reading skill, and parental income by including these variables as covariates in our ComBat model. This ensures that potential biological effects due to imbalances in these covariates across the scanner sites are not lost during the harmonization process. We employed the neuroCombat_sklearn library in the present analysis to perform ComBat harmonization on the data and remove any scanner or site related variation from the neuroimaging data.

2.3. PING

Data from the Pediatric Imaging, Neurocognition and Genetics (PING) Study database (http://ping.chd.ucsd.edu/) was used in the preparation of this article. PING was launched in 2009 by the National Institute on Drug Abuse (NIDA) and the Eunice Kennedy Shriver National Institute Of Child Health & Human Development (NICHD) as a 2-year project of the American Recovery and Reinvestment Act. The primary goal of PING has been to create a data resource of highly standardized and carefully curated magnetic resonance imaging (MRI) data, comprehensive genotyping data, and developmental and neuropsychological assessments for a large cohort of developing children aged 3 to 20 years. The scientific aim of the project is, by openly sharing these data, to amplify the power and productivity of investigations of healthy and disordered development in children, and to increase understanding of the origins of variation in neurobehavioral phenotypes. For up-to-date information, see http://ping.chd.ucsd.edu/.

The PING dataset consists of neuroimaging and behavioral data from 1119 participants between 3 and 20 years old. The imaging protocols and processing steps are outlined in Jernigan et al (Jernigan et al., 2016). In the present analysis, we used the diffusion tensor imaging (DTI) data provided by the authors of the PING study, which was processed using AtlasTrack (Hagler et al., 2009). This processed diffusion data includes the mean FA and MD values in 37 major white matter tracts, as well as the mean FA and MD values for the left hemisphere, right hemisphere, and the whole brain. In addition to the neuroimaging data, the PING dataset also provides behavioral assessments from the NIH Toolbox (Gershon et al., 2013a, 2013b), which includes an Oral Reading Recognition Test for ages 3 and above. We looked at age standardized scores on this reading assessment and classified individuals who scored below one standard deviation from the mean as struggling readers. This dataset also includes parental income, which we used as an index of socioeconomic status.

2.4. ABCD

The ABCD dataset consists of neuroimaging and behavioral data from the first observation of a ten-year longitudinal study. This time point includes data from 11,080 participants between 8 and 11 years old. The imaging protocols and processing steps are outlined in Casey et al. (Casey et al., 2018). In the present analysis, we used the minimally processed diffusion MR data provided in ABCD annual release 4.0 (available on the NIMH Data Archive). Like the HBN data, the ABCD data were processed using pyAFQ, which resulted in tract profiles for 24 major white matter pathways. In addition to the neuroimaging data, the ABCD dataset also provides the same age adjusted NIH toolbox Oral Reading Recognition Test (Gershon et al., 2013a, 2013b) found in the PING dataset. We looked at age standardized scores on this reading assessment and classified individuals who scored below one standard deviation from the mean as struggling readers. The ABCD dataset also includes many rich demographic and socioeconomic measures, some of which are not found in the other datasets, including parental income, neighborhood deprivation scores, and school achievement.

2.5. HCP-YA

The HCP-YA dataset (Harms et al., 2018) consists of neuroimaging and behavioral data from 1200 healthy young adults between the ages of 22 and 35. As with the HBN dataset, we used pyAFQ to perform tractometry on these data. For this analysis, specified parameters that yielded tract profiles for 24 major white matter pathways and four diffusion properties, FA, MD, MK (mean kurtosis), and AWF (axonal water fraction) for each tract. As a behavioral measure, we used the age adjusted NIH Toolbox Oral Reading Recognition Test to explore the relationship between the white matter and reading skill in this sample.

2.6. Quality control of neuroimaging data

To ensure that we did not include data from participants with low quality neuroimaging data, we relied on different quality control procedures depending on the dataset. For the HBN dataset, we made use of the quality control scores generated by Richie-Halford et al. (2022) and excluded individuals from the analysis who did not receive a sufficiently high quality control rating (Richie-Halford et al., 2022b). In the case of the PLING, PING and ABCD datasets, the raw and processed data both undergo quality control assessment before they are made available. The quality control procedures for the PLING, PING and ABCD datasets are outlined in Wierenga et al. (2018), Jernigan et al. (2016) and Hagler et al. (2019), respectively (Jernigan et al., 2016; Wierenga et al., 2018; Hagler et al., 2019). In the case of the HCP-YA data, we relied on the extensive QC procedures implemented in HCP (Marcus et al., 2013; Hodge et al., 2016).

2.7. Tract-wise analysis of the relationship between reading and math in HBN, PING, and ABCD

In our initial analyses of the white matter properties in poor and skilled readers, we first conducted group-wise comparisons of FA using a t-test at each node along the length of the left arcuate, ILF, SLF, and frontal and posterior components of the corpus callosum, as these tracts have been implicated in reading skill in the past (Wandell and Yeatman, 2013). To examine the anatomical specificity of these relationships, we also conducted the same analyses in the remaining 20 tracts as a control measure. After these initial comparisons, we then fit a linear-mixed effects model at each node in these same tracts predicting FA from reading skill (skilled or poor), age, measures of MRI data quality, which included mean framewise displacement and neighborhood correlation. Group differences were assessed by examining the FDR-corrected p-value on the beta-coefficient for reading skill.

We then looked to replicate these analyses using both the ABCD and PING datasets. For these replications, we limited our analyses to our initial tracts of interest. Because the ABCD data was processed with pyAFQ, the analysis of group differences for these data was identical to that used with the HBN data. However, because we only had access to the AtlasTrack derivatives of the PING sample, we could only conduct group comparisons of mean FA across our tracts of interest.

2.8. High-dimensional modeling of the relationship between reading and white matter in HBN, PING, ABCD, and HCP-YA

To generate reading score predictions from diffusion properties of the white matter, we fit a series of gradient-boosted random forest

models (XGBoost (Chen and Guestrin, 2016)). For each dataset, we fit three XGBoost models using either demographic information, white matter properties, or both demographic information and white matter properties as predictor variables. For all three datasets, we used age, SES, and, if available, geographic location and MRI data quality as predictors. The white matter predictors we used depended on the dataset and the tractometry software used to calculate diffusion properties. For the HBN, ABCD, and HCP-YA datasets, we used FA, MD, AWF, and MK in the 24 tracts identified by pyAFQ, while in the PING dataset we used the mean FA and MD values in the 37 tracts identified by AtlasTrack. Before being entered into the XGBoost models, each predictor variable was demeaned and scaled to unit variance. For each model, hyperparameter tuning was performed using a Bayesian optimization procedure (Louppe and Kumar, 2016) that samples over the hyperparameter space and performs 5-fold cross-validation to identify the best model fit. The final model fits were assessed using a held out test set that was counter-balanced with the training data to have similar demographic properties.

3. Results

3.1. A dynamic relationship between changes in the white matter and growth in reading skill

Past longitudinal and intervention studies have suggested that there is not a stable relationship between white matter properties and reading skill but, rather, that diffusion measures of the white matter and reading ability both change dramatically depending on an individual's educational environment (Yeatman et al., 2012a; Wang et al., 2017; Huber et al., 2018). To test the hypothesis that the white matter is part of a dynamic, experience-dependent system that is linked with learning over time, we used a large, longitudinal sample to model the relationship between changes in white matter diffusion properties and growth in reading skills. Here we define reading skills as single word reading and rely on different assessments, depending on the dataset, that largely tap into the same latent construct. These assessments have been used interchangeably across past studies investigating the relationship between white matter and reading skill (Supplemental Table 1). Based on past findings, we focused on the left arcuate, left SLF, left ILF, and corpus callosum, since these tracts are typically considered to be part of the core reading circuitry (Ben-Shachar et al., 2007; Wandell and Yeatman, 2013).

3.2. Development of the left arcuate tracks reading development

We first tested the hypothesis that changes in white matter properties in the left arcuate are linked with changes in reading scores using data from the Pediatric Longitudinal Imaging, Neurocognition, and Genetics study (PLING (Wierenga et al., 2018)). Based on the available diffusion MRI derivatives in this dataset, diffusion properties of the white matter were calculated using AtlasTrack (Hagler et al., 2009), which provides mean FA and MD (mean diffusivity) for 37 white matter tracts. We began this analysis by determining whether FA development follows a linear or non-linear trajectory in this sample. To do so, we constructed a linear mixed-effects model predicting FA using age as a fixed-effect and participant as a random effect. We then constructed a non-linear model by adding a quadratic term on age to the linear model. Wald tests comparing the two models revealed that the non-linear model did not fit the data better than the linear model (χ^2 (1) = 2.104, p = 0.147).

To explore the relationship between reading and white matter in the PLING sample, we fit a longitudinal linear mixed-effects model to predict mean-centered TOWRE (Test of Reading Word Efficiency) reading scores using time point, initial age, and mean-centered FA in the left arcuate, and mean FA in the left arcuate as fixed-effects and participant as random effects. TOWRE was used as our main reading outcome, as it was the only measure of reading available in this dataset. The inclusion of both mean-centered FA at each time point and overall mean FA allowed us to separate year to year change within an individual from mean FA differences between participants.

This model revealed significant effects of time point (t(244) = 14.141, p < 0.001), and mean-centered FA in the left arcuate (t(244) = 3.440, p = 0.0006). The effect of mean-centered FA suggests that, within an individual, changes in the diffusion properties of the left arcuate are linked with changes in reading scores across time, even after controlling for age-related increases in FA and overall FA level (Fig. 1B). Interestingly, we find no stable relationship between reading skill and white matter properties when we examine the correlation between TOWRE scores and FA in the left arcuate at each time point separately (Supplemental Fig. 1). Taken together, these results suggest that within individual changes in diffusion properties of the left arcuate fasciculus track gains in learning over time and call into question the notion of stable individual differences in white matter structure.

To estimate growth rates for FA, we fit a linear regression model within each individual, predicting each participant's FA value as a function of age in years (similar to Yeatman et al (Yeatman et al., 2012a).). The coefficients of the linear model serve as an estimate of the rate of change in FA over time within a given white matter tract. We applied a similar approach to generate individual estimates of reading score change over time.

We first examined the correlation between the rate of FA development in the left arcuate and the rate of raw reading score improvement and found a significant relationship between the two (r = 0.27, adjusted p = 0.02, Fig. 1B). Thus, children with more rapid growth in reading abilities also show more rapid white matter development in the left arcuate. We did not observe any relationship between the rate of FA development and rate of reading score development in any of the other tracts of interest (Fig. 1C). We next used the individual growth rates to examine the relationship between the rate of FA development and initial age to see if FA growth rates in the left arcuate differed across developmental stages. This correlation was not significant (r = −0.210, p = 0.08), suggesting that the rate of FA change in the left arcuate is not significantly different from linear over the age range present in the sample.

3.3. Reading skills predict future FA development in the left arcuate

Given the relationship between changes in FA in the left arcuate and changes in reading skill, we then tested whether changes in one variable preceded changes in the other or if these changes occurred in parallel. To model the longitudinal interplay between white matter development and reading, we constructed a multi-level vector autoregression model (mlVAR (Epskamp et al., 2018a); Bringmann et al., 2013). Within this model, we used the time series of FA measurements in the left arcuate and TOWRE reading scores to assess whether reading scores at a given time point are explained by FA at the previous time point or vice versa. This analysis revealed that FA in the left arcuate at one time point does not predict reading scores at the next time point (t(162) = –0.01, p = 0.971), but that reading skill at one time point does, in fact, predict FA at the following time point (t(162) = 5.44, p < 0.0001).

To compare the beta-coefficients of the model path connecting past reading scores with future FA against the path linking past FA with future reading scores, we conducted a bootstrapped difference test (Epskamp et al., 2018b). To do so, we generated a bootstrap distribution (n = 2000) of mlVAR coefficients and constructed a 95% confidence interval around the difference between the reading to FA coefficient and the FA to reading coefficient. The interval of the difference between the two coefficients did not include 0 (Bootstrapped 95% CI: [0.728, 0.748]) suggesting that the two coefficients are significantly different. Thus, growth in reading skills from one year to the next predicts future changes in the white matter, whereas developmental changes in the white matter do not predict future gains in reading skill.

The mlVAR model revealed that reading skill at one time point predicts FA in the left arcuate at the next, however, this model does not

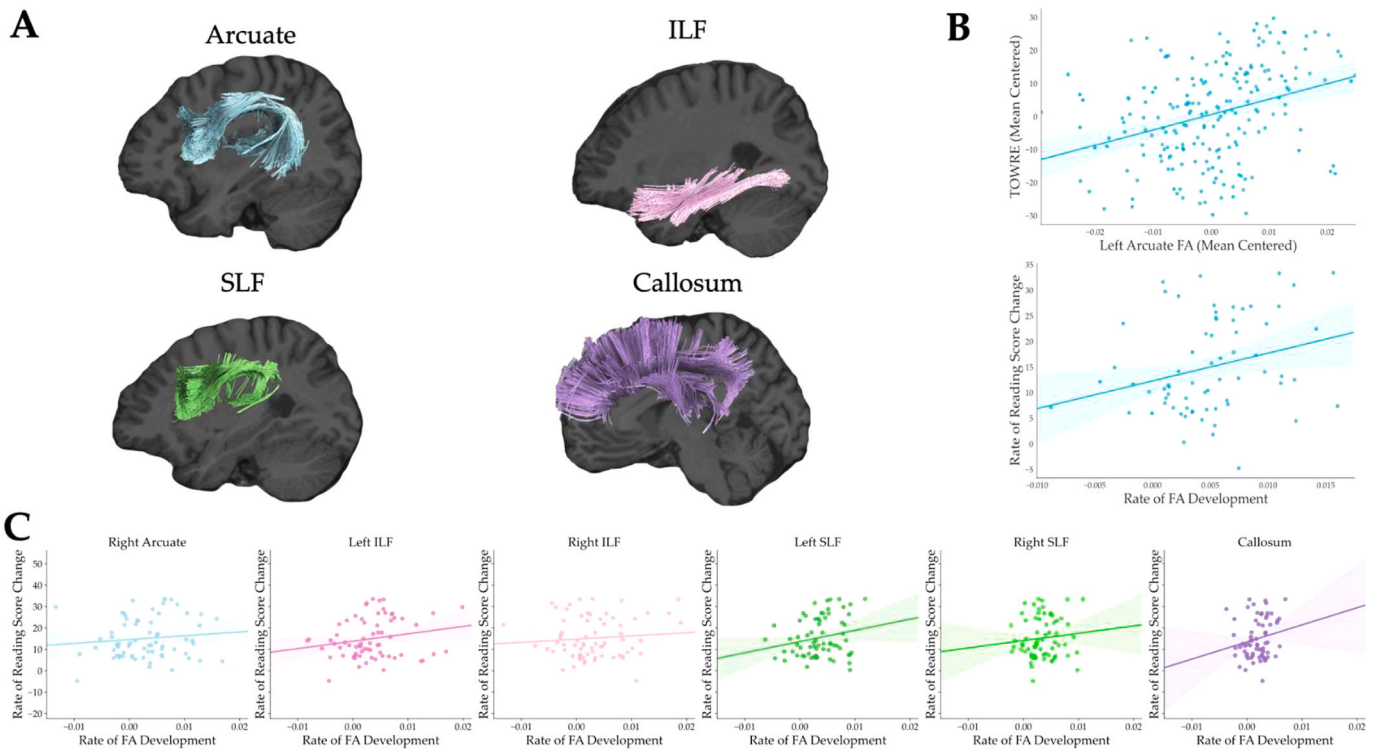


Fig. 1. **A**: Rendering of the arcuate, inferior longitudinal fasciculus, superior longitudinal fasciculus, and corpus callosum identified with AtlasTrack. Due to constraints with the PLING dataset, only renderings of the right arcuate, ILF, and SLF are available. **B.** Upper Panel: Mean-centered FA (Fractional Anisotropy) values at each time point for the left arcuate correlate with mean centered reading scores assessed at each time point, illustrating the longitudinal relationship between changes in FA and changes in reading skill. Lower Panel: Relationship between individual rates of FA development in the left arcuate and rate of raw TOWRE (Test of Reading Word Efficiency) reading score change. **C**: Relationship between individual rates of FA development in the right Arcuate, left and right ILF, left and right SLF, and callosum and rate of raw TOWRE reading score change in the longitudinal PLING dataset.

provide insight into the relationship between the rate of reading development and the rate of FA development in the left arcuate. To better understand this dynamic, we constructed a parallel process latent growth curve model of the longitudinal growth of reading skill, FA, and the relationship between the two. Because the model did not converge properly due to an insufficient number of participants at the fifth time point, only observations from the first four time points were included in this analysis. Fig. 2 presents a path diagram illustrating the hypothesized growth trajectories of reading skill and FA in the left arcuate.

To assess the impact of change in reading score on change in FA (and vice versa), we generated two models incorporating two different regressors (in place of covariance structures): one predicting the slope of FA from the slope of reading and another predicting the slope of reading from the slope of FA. Based on modeling of individual growth rates presented above, we defined this latent growth curve using linear growth rates for both reading and FA. The overall model fits were acceptable ($\chi 2$ (Deutsch et al., 2005) $= 33.41$, $p = 0.056$, RMSEA $= 0.085$, TLI $= 0.973$). Additionally, the regression predicting FA slope from reading slope was significant ($z = 1.956$, $p = 0.05$), whereas the regression predicting reading slope from FA slope was not ($z = -1.311$, $p = 0.190$). Taken together, the results from these models suggest that, while both FA and reading skill develop over time, the rate of individual

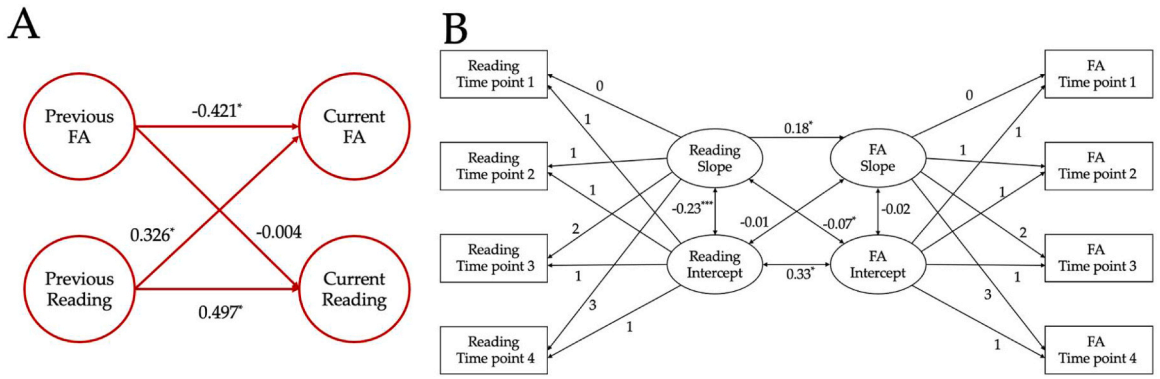


Fig. 2. **A.** Path diagram illustrating the mlVAR model capturing the temporal dynamics between the development of reading and FA in the left arcuate over time. The values represent the beta-weights associated with each path within the model. **B.** Path diagram outlining the parallel process latent growth curve model capturing the interplay between longitudinal growth in reading and FA development in the left arcuate. Rectangles represent observed reading or FA values at a given time point and ovals represent latent slope and intercept variables. Unidirectional arrows represent regressions within the model and the adjacent numbers signify the associated beta-coefficients. Bidirectional arrows represent covariance structures and the adjacent numbers signify the covariance values.

reading gains predicts the rate of future FA development, while the rate of FA development does not predict the rate of reading gains.

3.4. Rate of FA development in the right arcuate and left ILF are not related to changes in reading

To test whether the same longitudinal relationships existed between reading and additional white matter tracts, we first constructed the same longitudinal mixed-effects model in the right arcuate, a tract that has not been implicated in the literature as a core component of the reading circuitry (Ben-Shachar et al., 2007; Wandell and Yeatman, 2013). This model revealed a significant main effect of time point ($t(244) = 10.811$, $p < 0.00001$) but no effect of FA in the right arcuate ($t(244) = 1.325$, $p = 0.186$) nor initial age ($t(244) = 0.530$, $p = 0.597$). We also conducted the same longitudinal mixed-effect analysis using FA in the left ILF since past findings have also identified this tract as part of the reading circuitry. Again this model revealed a significant main effect of time point ($t(244) = 13.998$, $p < 0.00001$). However, we did not observe significant main effects for FA in the left ILF ($t(244) = 1.526$, $p = 0.128$) nor initial age ($t(244) = -0.884$, $p = 0.377$). These main effects suggest that neither FA in the right arcuate nor the left ILF significantly relate to changes in reading skills over time. There was also no significant relationship between FA in right arcuate or ILF and reading scores when the time points were examined separately.

3.5. No static relationships between white matter properties and reading ability

After establishing the dynamic, longitudinal relationship between white matter development and reading development, we then attempted to replicate past findings from the literature which have suggested that static differences in the white matter explain individual differences in reading skill. Using the HBN dataset (Alexander et al., 2017), one of the most diverse large-scale pediatric samples to date, we compared the white matter diffusion properties of struggling readers (again using a typical cutoff for dyslexia) and a control group to test the hypothesis that stable individual differences in white matter properties relate to reading skill, as measured by age-standardize TOWRE composite scores.

The HBN data was processed with QSIPrep (Cieslak et al., 2020), rigorous quality control was applied to each participant's data (Richie-Halford et al., 2022a), and tractometry was performed with pyAFQ (Kruper et al., 2021). After identifying 24 major white matter tracts, pyAFQ samples each tract to 100 nodes and provides various diffusion metrics for each node (see Methods for overview). The following node-wise group comparisons focused on the left arcuate, left ILF, left SLF, and corpus callosum, which have been most consistently implicated in past studies examining the relationship between the white matter and reading skill (Ben-Shachar et al., 2007; Wandell and Yeatman, 2013). Based on past cross-sectional findings, the expectation is that children with low reading scores will show reduced FA values in the left arcuate, left ILF and left SLF compared to the control group and increased FA in the posterior callosum. Additionally, we analyzed 20 control tracts to examine anatomical specificity of reading effects in the large and diverse HBN sample (Supplemental Fig. 2).

An initial group comparison using minimal quality control and not controlling for age or SES revealed reduced FA in areas of the right SLF and the callosal motor fibers in struggling readers (corrected $p < 0.05$, Supplemental Fig. 3), but not in left arcuate, left ILF, or left SLF (all corrected $p > 0.05$). After filtering for quality control (See Methods), performing ComBat harmonization on the tract profile data to remove the effects of the scan site, and controlling for various confounds, a group comparison at each node failed to reveal any significant differences in FA (Fig. 3, all adjusted $p > 0.05$). Fig. 3 shows the differences in FA between the two reading groups at each node after controlling for age, SES, MRI data quality, and geographic location. There were no nodes with significant group differences. Since the HBN data contains a

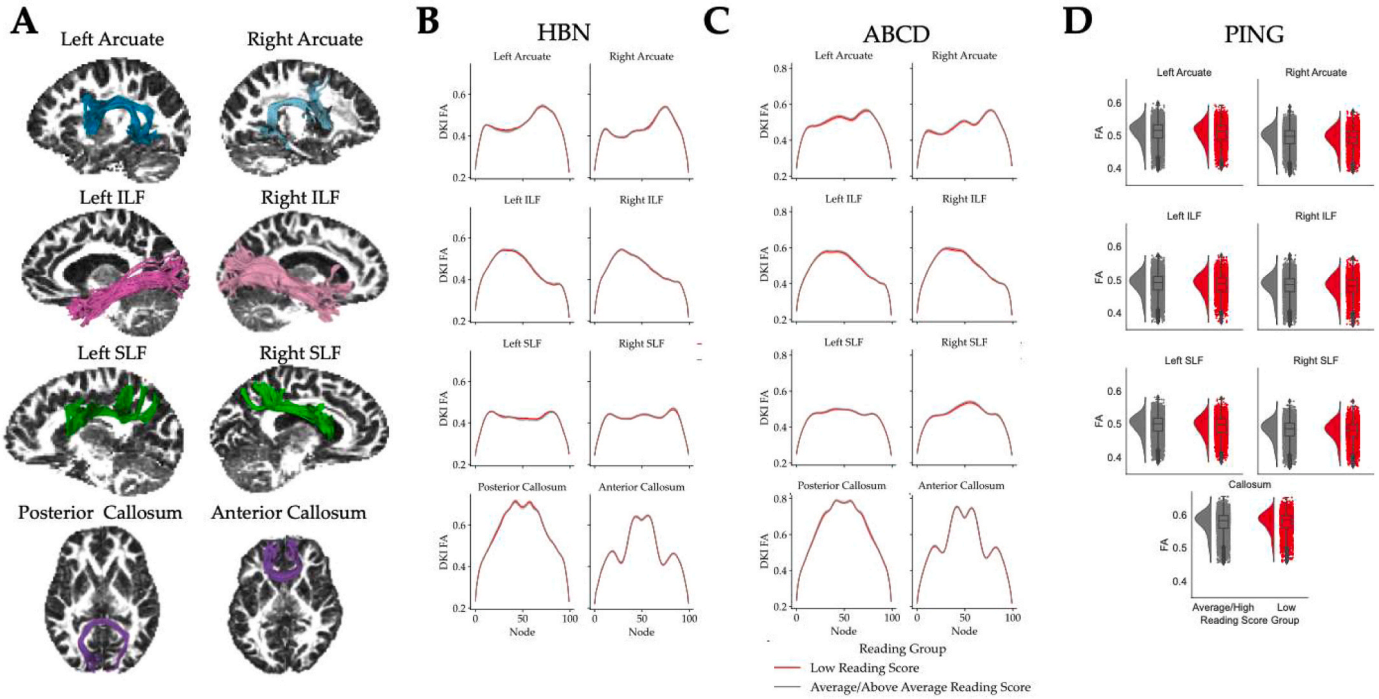


Fig. 3. **A:** Rendering of the eight tracts of interest identified with pyAFQ overlaid on the T1-weighted image of the corresponding hemisphere. Note: the posterior and anterior callosum are included as separate tracts though both are part of the corpus callosum. **B**: FA (Fractional Anisotropy) profiles for the core reading circuitry in the HBN data derived by the default pyAFQ pipeline for reading groups after controlling for age and harmonizing on scanner site. The shaded area around each line represents one standard error. **C**: FA profiles for the core reading circuitry in the ABCD. **D**: Rain cloud plots showing the distributions of mean FA values for the core reading circuitry by reading group in the PING data set. There are no significant differences in the diffusion properties between reading score groups across the two datasets. There is a negligible effect of reading group on FA values across all seven tracts.

broad distribution of ages, we repeated this group comparison separately for three age bins as roughly found in the literature: 5–9 years old (Yeatman et al., 2012a; Wang et al., 2017; Huber et al., 2018; Niogi and McCandliss, 2006b), 10–15 years old (Keller and Just, 2009; Hasan et al., 2012; Yeatman et al., 2011), 14–21 years old (Steinbrink et al., 2008). As with the full sample, we found no group differences across any of the age ranges (Supplemental Figure 7).

To validate the surprising lack of group differences observed in the HBN dataset, we also explored the relationship between white matter and reading in the PING dataset (Alexander et al., 2017; Jernigan et al., 2016; Casey et al., 2018). In the PING dataset, the NIH Toolbox Oral Reading Recognition Test is the only available measure of reading skill. As with the PLING data, diffusion metrics for PING data are provided based on AtlasTrack (as opposed to pyAFQ; see Methods for overview). Therefore, we could only examine mean FA and MD for 37 white matter tracts in this analysis. In these data, we looked to see if we could detect differences in the mean FA across the tracts that comprise the core reading circuitry. After controlling for age and parental income, we again did not find any significant differences in mean FA between the struggling readers and control groups (Fig. 3, all $p > 0.05$).

After observing no group differences in the HBN and PING datasets, we turned to the ABCD dataset (Alexander et al., 2017; Jernigan et al., 2016; Casey et al., 2018), the largest ongoing developmental neuroimaging study. Similar to the HBN dataset, tractometry was performed using pyAFQ. We constructed a series of regression models to explore the extent to which reading skills, as measured by the NIH toolbox Oral Reading Recognition Test, predict FA in the core reading circuitry. In these models, we leveraged the rich demographic data only present in the ABCD dataset, including parental income, neighborhood deprivation, and school achievement to create a fuller representation of socioeconomic status. After controlling for these demographic factors, the addition of reading group to the model had a negligible effect on the amount of variance explained between the two models (all $\Delta R^2 < 0.003$, all all Cohen's $f^2 < 0.003$, Fig. 3).

3.6. Diffusion properties fail to predict reading scores above demographics in three large-scale datasets

After failing to observe any significant differences between individuals with low and high reading scores, we explored the idea that static white matter differences related to reading skill are not localized to a single tract but, rather, form a distributed network that might have a nonlinear relationship to reading ability. To test this hypothesis, we fit a series of gradient-boosted random forest models (XGBoost (Chen and Guestrin, 2016)) to determine how well diffusion properties from the entire white matter, not just the reading circuitry, serve to predict reading scores. For this particular analysis, we prioritized predictive accuracy (as opposed to interpretability) (Shmueli, 2010). We chose the XGBoost algorithm because of its exceptional performance across a wide range of machine learning applications (Parsa et al., 2020; Ogunleye and Wang, 2020; Gumus and Kiran, 2017) and its ability to capture complex, non-linear relationships that would be missed by a linear regression model.

We began this analysis by fitting a series of models to predict TOWRE scores from a range of factors in the HBN dataset. Our first model contained only age, SES, and geographic location as predictor variables; the second model contained demographic variables and measures of MRI data quality; the third model contained diffusion properties (FA, MD, AWF, MK) from all 24 tracts delineated in the dataset and measures of MRI data quality; the fourth model contained white matter properties and measures of MRI data quality; the full model contained demographic variables, white matter properties and measures of MRI data quality as predictors. Each predictor was first demeaned and then scaled to unit variance before being entered into the XGBoost models. The performance of these five models is presented in Supplemental Figure 9.

Cross-validated model R^2 showed that the demographics-only and white-matter-only models explain roughly the same amount of variance in reading scores in the HBN data (Fig. 4). The combination of white matter features and demographic information as predictors did not improve model fit above the models with only demographic information or only white matter features. Thus, as expected, differences in demographics (e.g., SES) did explain some variance in reading scores (Jensen et al., 2005; Henriques et al., 2021; Fieremans et al., 2011). The model with only diffusion properties did predict roughly 3.5% of the variance in reading scores (Table 2). However, white matter properties did not explain additional variance in reading ability above and beyond demographics, indicating that there was not a specific relationship between reading skill and the white matter, beyond what is predicted by sociodemographic factors (Table 2). Furthermore, when we divided the HBN sample into low, medium, and high SES bins and applied the same modeling pipeline, neither demographic features nor white matter properties explained any variance in reading skill (Table 2). We then turned to a different measure of reading ability, the Wechsler Individual Achievement Test (WIAT), to see if white matter properties could predict a different measure of reading other than TOWRE. Similar to our first set of models, diffusion properties of the white matter did not explain additional variance in reading skill above and beyond demographic factors (Supplemental Fig. 4).

To further examine the surprising finding that individual differences in the white matter do not, in fact, predict reading abilities (above and beyond demographic factors), we reprocessed the data from scratch using a different approach to tractography. Specifically, we recalculated tractometry properties using tractography data generated by QSIPrep (Cieslak et al., 2021) using anatomically constrained tractography (Smith et al., 2012) (ACT; see Methods) to see if the lack of reading ability prediction stemmed from our choice of tractography parameters during the initial processing pipeline. These new tractometry properties again revealed group differences in the tract profiles before quality control, but these differences disappeared when we applied more rigorous quality control and harmonized the data across scanner sites (Supplemental Figure 5). We then conducted the same XGBoost modeling approach with these data and again found that white matter properties did not predict reading scores above and beyond demographic information (Supplemental Figure 6).

Finally, we used the AtlasTrack derivatives present in the PING dataset and the pyAFQ outputs generated from the ABCD dataset to train a series of XGBoost models to predict reading scores from either demographic factors, diffusion properties of the white matter, or both demographic factors and white matter in these independent datasets. The models trained on the PING data showed that demographic features, namely socioeconomic status, predicted reading scores (cross-validated $R^2 = 0.040$) and that the addition white matter features to the model did not increase predictive power (cross-validated R^2 =0.045, t (Huber et al., 2021)= 0.383, $p = 0.36$, Fig. 4C). The models trained on the ABCD data also revealed that demographic features explained roughly 10% of the variance in reading scores (cross-validated $R^2 = 0.101$), and white matter features did not improve predictive power of the model (cross-validated R^2 =0.078).

3.7. Maturational differences in white matter properties predict age

To ensure that the lack of cross-sectional effects did not reflect poor data quality or modeling errors, we sought to benchmark our models against a finding with a well-established effect size. The largest and most consistent effect in the diffusion MRI literature is that FA increases and MD decreases with age (Yeatman et al., 2014; Lebel et al., 2008; Eluvathingal et al., 2007; Lebel et al., 2019). Based on this, we would expect to be able to explain a significant portion of the variance in age in the present sample using diffusion data. To evaluate the signal quality of the diffusion data, we fit models to predict each participant's age using the same XGBoost models. As expected, the demographics-only model failed to predict age above chance confirming that there was not a sampling

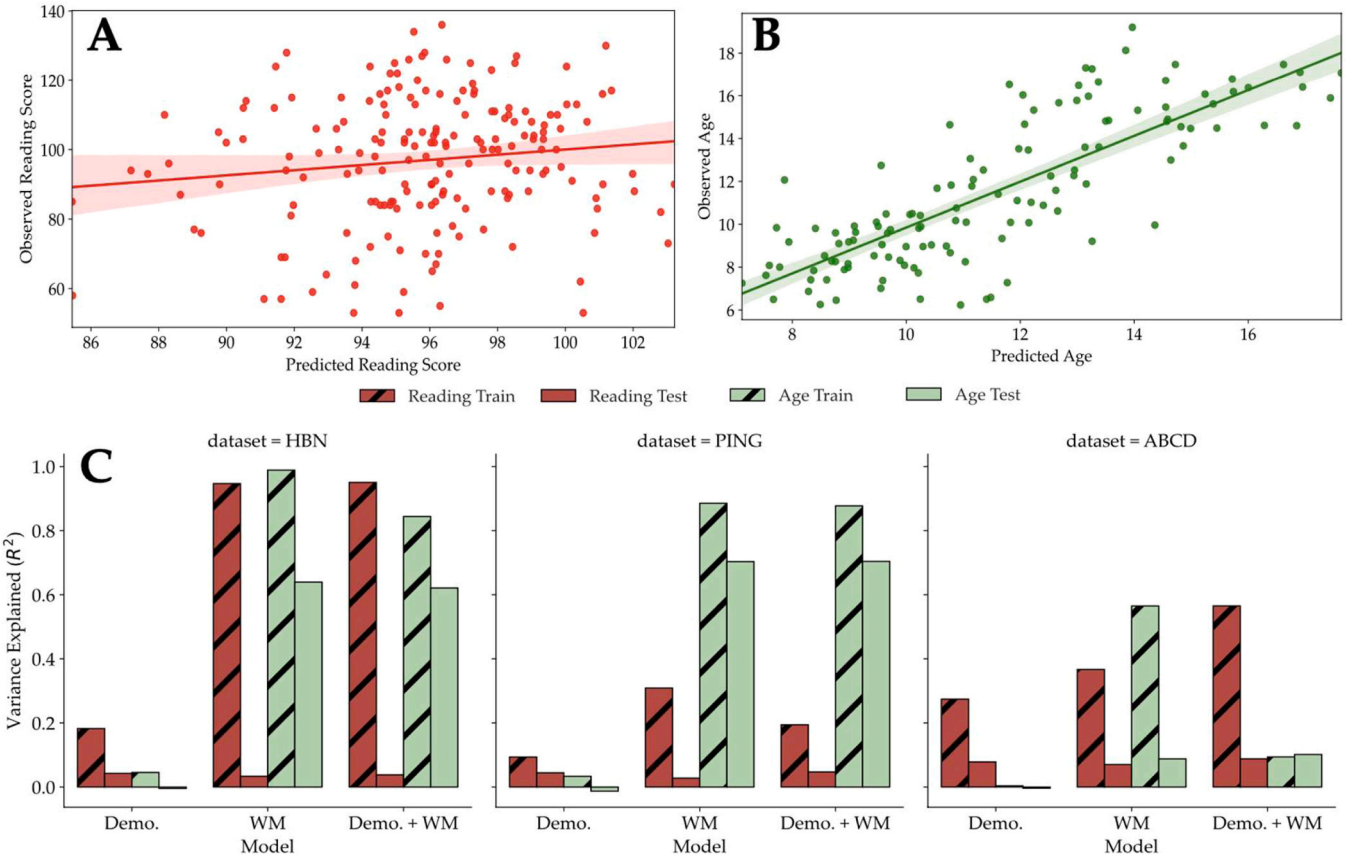


Fig. 4. A. Predicted vs observed reading scores of the test data generated by the XGBoost model trained on white matter features from the HBN data (R^2 =0.059). **B.** Predicted vs. observed age from XGBoost model trained on white matter features from HBN data (R^2 =0.64). **C.** Train and test R^2 scores for XGBoost models predicting reading (red) and age (green) in the HBN (left panel) and PING (center panel) and ABCD (right panel) datasets.

bias. The white matter model, on the other hand, predicted 64% of the variance in age in the HBN dataset and 70% of the variance in age in the PING dataset (Fig. 4C). These predictions are consistent with the state-of-the-art in "brain age" calculations (Brown et al., 2012) confirming, first, that the diffusion MRI data was sufficiently high quality to accurately index white matter maturation and, second, that our modeling approach was able to accurately capture these effects.

In contrast to the HBN and PING datasets, white matter properties only predicted about 7% of the variance in age in the ABCD dataset. However, all the participants in the ABCD sample are within two years of age of each other and, therefore, there is not much maturational variation in the white matter properties for the model to learn.

3.8. *Reading score - white matter relationships are sample-dependent*

The present analysis leveraged the largest diffusion MRI sample to date to explore the relationship between white matter and reading abilities. Contrary to the compendium of small, single-observation studies, this large sample did not reveal any stable predictors of reading ability in the white matter. One potential explanation for this discrepancy is that previous studies were picking up on real effects that reflected specific characteristics of the small, homogenous samples that do not generalize to the population at large. The HBN dataset is an extremely neurodiverse sample of children from a variety of demographic backgrounds, with 87% of participants having at least one diagnosis and 53% of participants having more than one diagnosis.

To test the hypothesis that seemingly conflicting results across past studies stem from specific features of the participant recruitment strategies, we tested whether we could predict reading scores using a small, homogeneous sub-sample of the entire HBN dataset that excluded participants based on strict quality control measures and co-occurring diagnoses. We began by selecting all the participants between the ages of 8 and 12 years old from one site (CBIC) and excluding any participants diagnosed with autism, anxiety, or ADHD, as these diagnoses have been shown to correlate with reading skills in different ways (Ostrolenk et al., 2017; Germanò et al., 2010; Grills-Taquechel et al., 2012). Within this limited sub-sample, we then identified the 40 top and bottom readers based on their TOWRE scores to create groups that differed dramatically in terms of reading ability. The resulting, single-site sample was skewed towards high SES participants, as is the case with many single-laboratory studies. We split this sub-sample into a training set of 60 participants and a test set of 20 participants. We then trained an XGBoost classifier on the train set using 5-fold cross-validation before predicting reading level (high or low) in the test set. The classifier trained on white matter properties had a classification accuracy of 0.714 in the held-out test sample. In addition to this classification model, we trained an XGBoost model to predict reading scores in this same sub-sample, to facilitate comparison across our other modeling pipelines. The model trained on white matter data to predict reading scores in this sub-sample predicts roughly 11% of the variance in reading scores in the test set (Table 2). However, when we thresholded the reading score predictions made by our regression model using the maximum score for the low reading group as a cutoff, we found that the model had a classification accuracy of 0.52 in the test sample, just slightly above chance performance.

Next, we examined the relationship between white matter and reading ability in a large-scale, dataset consisting of healthy, college-aged young adults (HCP). The original HCP recruitment screened for many different developmental, neurological and psychiatric disorders and reflects a more homogenous sample than, for example, the HBN data (Van Essen et al., 2012; Chin Fatt et al., 2021) and is substantially less

Table 2
Cross-validated R^2 test scores for TOWRE reading score predictions made from XGBoost models trained on white matter properties. The diffusion properties used to train the models vary due to differences in the tractography software used on the various data sets.

Dataset	Diffusion Predictors	Number of Prediction Features	Cross-validated R^2 Score (Test Set)
HBN - Full Sample	FA, MD, MK, AWF	9600	0.033
HBN - Full Sample	FA	2400	< 0.001
HBN - Full Sample	MD	2400	-0.032
HBN - Full Sample	MK	2400	-0.019
HBN - Full Sample	AWF	2400	0.010
HBN - Full Sample	Mean FA	24	0.014
HBN - Full Sample	Mean MD	24	-0.021
HBN - Full Sample	Mean MK	24	-0.001
HBN - Full Sample	Mean AWF	24	0.015
HBN - Low SES	FA, MD, MK, AWF	9600	-0.001
HBN - Medium SES	FA, MD, MK, AWF	9600	-0.035
HBN - High SES	FA, MD, MK, AWF	9600	-0.126
HBN - Site CBIC	FA, MD, MK, AWF	9600	-0.076
HBN - Site CUNY	FA, MD, MK, AWF	9600	-0.844
HBN - Site RU	FA, MD, MK, AWF	9600	-0.016
HBN - Homogenous subset	FA, MD, MK, AWF	9600	0.114
HCP-YA - Top/ Bottom 100 Readers	FA, MD, MK, AWF	9600	0.079
PING - Homogenous subset	Mean FA, Mean MD	37	-0.044

diverse than the US population. Using an XGBoost model, we found that diffusion properties of the white matter calculated by pyAFQ were able to predict 7.9% of the variance in reading scores, as measured by the NIH Toolbox, in the top and bottom 100 readers, in line with other predictions of cognitive ability made using the HCP dataset (Rasero et al., 2021). Thus, in these two cases representing samples that were a) more demographically homogeneous, b) differed substantially in terms of reading ability and, arguably c) are more similar to a typical sample collected by a single university research group, white matter properties did predict individual differences in reading ability.

However, using the same approach in the PING dataset, we were unable to predict reading scores from diffusion properties. Similar to the HBN data, we generated this subsample by taking the 40 highest vs lowest scoring readers between the ages of 8 and 12 years old. We again trained XGBoost models on the data from this subset and were unable to predict reading scores from the average diffusion properties of the white matter (Table 2). The lack of prediction in the PING subset suggests that mean diffusion properties may not be suitable for generating cross-sectional behavioral predictions, even in homogenous samples.

4. Discussion and conclusions

We explored two hypotheses surrounding the relationship between white matter and reading skills: a) that white matter and reading skill are dynamically linked over time and change with factors such as an individual's educational environment and b) that white matter properties act as static biomarkers that predict differences in reading abilities. To test the first hypothesis, we examined a five-year longitudinal dataset, which revealed that individual growth rates in the left arcuate are linked with growth in reading scores over time. Further modeling revealed that gains in reading predict future changes in FA in the left arcuate fasciculus. These models suggest that individual differences in learning are related to white matter development rather than white matter differences serving as constraints on the learning process.

To explore the second hypothesis, we analyzed three large-scale cross-sectional datasets (HBN: n = 777, ages: 5–21; PING: n = 1119, ages: 3–20; ABCD: n = 11,080, ages: 8–11) and found no stable relationship between white matter diffusion properties and reading scores. To capitalize on the strengths of both explanatory and predictive modeling (Shmueli, 2010), we conducted group comparisons to explore the specific white matter properties underlying differences in reading skill and also leveraged high-dimensional models to predict reading scores from white matter properties. Across all three of these datasets, univariate group comparisons revealed no differences between poor and skilled readers and XGBoost models were unable to predict reading scores from diffusion features (after controlling for SES). However, analysis of a large-scale adult dataset (HCP-YA: n = 1200) and a subset of the HBN data revealed that diffusion properties of the white matter serve to predict reading scores in more homogenous samples, suggesting that group differences in the white matter between typical and struggling readers may depend on the makeup of the sample. All together, these results suggest that properties of the white matter may not necessarily serve as static traits that differentiate individuals but, rather, that white matter and reading skill may be part of a dynamically linked system that changes over time.

The longitudinal dynamics between reading skill and white matter properties observed in the PLING dataset serve to highlight the role that an individual's environment may play in driving the development of brain-behavior relationships. In these data, not only did changes in reading skill track changes in FA in the left arcuate, the rate of individual reading gains predicted increases in FA in the left arcuate, suggesting gains in reading precede, and potentially influence, changes in the white matter. Other longitudinal studies have shown that skilled readers demonstrate positive rates of FA development in the left arcuate and ILF, whereas poor readers demonstrate shallower developmental slopes in the same tracts (Yeatman et al., 2012a; Wang et al., 2017), but the present study is the first to examine temporal precedence of longitudinal relationships in the reading circuitry. Future work should expand these analyses to additional white matter tracts to better understand the longitudinal relationship between white matter and learning throughout the entire brain, not just within the reading circuitry.

Nevertheless, questions remain as to how growth opportunities within an individual's learning environment relate to changes in both academic skills and brain properties. In the PLING dataset, only three participants had an initial standardized TOWRE composite score below 85 making it impossible to study the differences in growth trajectories between struggling and typical readers. Additionally, the present sample includes a broad range of ages and we therefore cannot assess how these longitudinal dynamics change across different stages of development. It also bears mentioning that the demographic makeup of the PLING sample largely resembles that of the homogeneous HBN sub-sample and that these longitudinal findings may not necessarily generalize to a more demographically and socioeconomically diverse population. Future longitudinal and intervention studies will be critical for understanding these developmental dynamics and determining the relationships between demographic and environmental factors, learning, and white matter plasticity.

The longitudinal interplay between white matter growth and reading gains raises questions about previous reports of a static link between white matter structure and reading skills. Many studies have found differences in the white matter properties of dyslexic versus typical readers based on cross-sectional observations of small samples (Vandermosten et al., 2012; Saygin et al., 2013). However, the present analysis of the HBN, PING, and ABCD datasets (totaling nearly 13,000 participants) failed to reveal any meaningful group differences in the white matter properties of struggling and typical readers. Furthermore, state-of-the art machine learning models were unable to predict reading scores from white matter properties above and beyond demographic

factors. This observation is in line with recent work showing that the effect sizes of many brain-behavior relationships are much smaller (or non-existent) than estimates from small samples (Marek et al., 2022).

These results are also generally in line with those reported by Meisler and Gabrieli (Meisler and Gabrieli, 2022) which did not find any significant relationship between FA and reading skills in a subset of the HBN dataset. Though the authors observed a relationship between FA and phonemic decoding in the right SLF and the left ICP in participants above age 9, their analysis relied on a different analytic pipeline (TractSeg (Wasserthal et al., 2018)), which may explain the subtle differences between their findings and the present results. Regardless of the differences between these independent analyses of the HBN dataset, both sets of results suggest that the relationship between the white matter and reading skill may not necessarily be reducible to static, individual differences in the diffusion properties of the white matter. White matter properties are highly plastic and changes in the diffusion signal have been detected on the timescale of hours (Sagi et al., 2012) and months (Keller and Just, 2009; Mackey et al., 2012) after a learning experience. Therefore, it is possible that these datasets serve as a snapshot of a dynamic, experience-dependent system in flux and do not capture the longitudinal relationships between brain, behavior, and environment.

Based on this interpretation, however, one might expect to observe a more stable relationship between reading and white matter properties in adults. Although, when we divide the HBN sample into distinct age bins, we are unable to observe a relationship between reading skill and white matter properties, even in the oldest age range (Supplemental Figures 7 and 8), in the HCP-YA adult sample, we do observe a link between white matter properties and reading skill. Together, these results suggest that environmental and/or developmental factors may dynamically influence the relationship between reading and white matter over the course of development and lead to detectable group differences in adulthood. However, it could also be the case that there exist dynamic traits moderating the relationship between white matter and literacy that obscure group differences over the course of development but lead to detectable differences in adulthood. Future longitudinal studies will be necessary to better understand the dynamics between white matter plasticity and literacy, especially considering that white matter pathways continue to develop throughout childhood and into adulthood (Lebel et al., 2019).

The lack of meaningful group differences in the HBN, PING, and ABCD samples do not necessarily serve to dismiss past findings linking properties of the white matter to reading skill but, rather, present new challenges for the field consider in order to reconcile the seemingly conflicting findings from large-scale data sets, small single-lab studies, and meta-analyses. First, in the present analyses, we rely on traditional diffusion MRI measurements, namely FA. It could be the case that novel diffusion measurements might serve to capture static relationships between the white matter and literacy. Second, as highlighted in recent work using large-scale fMRI datasets, the effect size of various brain-behavior relationships decreases as a function of sample size and makeup (Marek et al., 2022). Moreover, models that are trained on a biased demographic group often do not generalize to more diverse samples (Li et al., 2022). It is possible that sampling procedures impact the relationship between diffusion properties of the white matter and academic skills and may explain the seemingly contradictory results between the present study and past findings.

One of the strengths of these large scale datasets, especially the HBN sample, is that they include participants from a wide range of geographic and demographic backgrounds, representing diversity that is rarely present in the samples collected by a single lab. Thus, these large-scale datasets provide a new opportunity to explore brain-behavior relationships while controlling for the robust relationship between clinical and socioeconomic factors and academic outcomes. For example, the HBN sample is also extremely neurodiverse: nearly all the individuals in the dataset have some sort of clinical diagnosis. In the final sample used in our analysis, only 35 of 777 participants (4.5%) had no diagnosis whatsoever. Some of these co-occurring diagnoses can impact reading scores and white matter properties and it could be the case that these multiple, overlapping clinical diagnoses may obscure potential brain-behavior relationships. For instance, many individuals diagnosed with autism have been shown to have hyperlexia (Ostrolenk et al., 2017), whereas ADHD or anxiety diagnoses have been linked to lower reading scores (Germanò et al., 2010; Grills-Taquechel et al., 2012). On the other hand, past findings from single laboratories that demonstrate a relationship between white matter properties and reading skill are typically based on a single observation of smaller, demographically homogeneous samples, usually recruited from the community surrounding a university. Additionally, these samples often exclude individuals with certain diagnoses. Inconsistent sampling procedures may explain why some studies report increased FA in struggling readers (Hasan et al., 2012; Yeatman et al., 2011) and others report reduced FA in struggling readers (Wang et al., 2017; Keller and Just, 2009; Niogi and McCandliss, 2006b).

Interestingly, when we generate a small ($n = 80$) homogenous subsample of the HBN data, excluding participants with autism, ADHD, or anxiety diagnoses, we find that high-dimensional models trained on diffusion properties of the white matter do predict some variance in reading scores. However, it should also be noted that this subset of the data is highly skewed in terms of socioeconomic status, with most of the participants coming from upper SES households. This finding parallels the results observed in the HCP-YA data, which also revealed that properties of the white matter served to predict reading scores in a large, relatively homogenous sample of adults. The recruitment procedures in many single-lab studies may lead to more discernible relationships between individual differences in white matter properties and reading skill but these results might not generalize beyond the sample. In fact, for individuals with high SES backgrounds, genetic factors have been shown to exert more of an influence on FA than those from lower SES backgrounds (Chiang et al., 2011), suggesting that in homogenous, high SES samples, genetic factors may influence the relationship between white matter and reading skill in a manner that may not replicate in more diverse study populations.

These findings also raise questions surrounding the relationship between SES (and other environmental factors more broadly), literacy, and white matter development. In the present analysis, the models trained exclusively on white matter features did predict some variance in reading skills. However, this variance largely overlapped with the variance explained by demographic factors, given that the models trained on both white matter and SES did not serve to explain additional variance in reading abilities and white matter did not predict reading skill whatsoever in models trained on data from distinct SES bins. Previous studies relying on smaller, more homogenous participant populations might have been picking up on this indirect relationship. Untangling the complex relationships between environmental factors such as SES, brain development, and academic achievement is a complex issue (Li et al., 2022) and researchers leveraging large-scale, single--observation datasets need to carefully consider how to incorporate covariates, such as clinical diagnoses and SES, into their analyses.

It also bears mention that these large-scale datasets generally consist of data collected using multiple scanners, often made by different manufacturers. Subtle differences across scanners may introduce an additional source of variability to the data despite efforts to harmonize scan sequences across sites (Alexander et al., 2017; Jernigan et al., 2016; Casey et al., 2018). We employed state-of-the-art approaches to harmonize the neuroimaging data across scanner sites (Fortin, 2017, 2018; Johnson et al., 2007) and reduce the impact of scanner site on the variability of the diffusion measures. However, when we generated predictive models from each scan site separately, only data from one site predicted any variance in reading scores. Thus it is possible that differences in scanners and pulse sequences may obscure subtle differences in the white matter that do relate to reading skill.

As the field of developmental cognitive neuroscience moves into the

era of Big Data with diverse groups of participants, longitudinal measurements will be of particular importance to answer questions surrounding the relationship between the brain, academic environment, and learning. Past meta-analyses have failed to identify reliable neurological profiles for individuals labeled as "learning disabled" (Moreau et al., 2018; Peters et al., 2018), suggesting that a more individualized approach to studying learning may be important, and densely sampled functional neuroimaging studies have found that an individual's neurobiology is often distinct from group-level patterns (Yip and Konova, 2021). A single snapshot of a dynamic system can be misleading and longitudinal studies are critical for capturing the within-individual interplay between environmental factors, brain development and learning.

Historically, the field of developmental cognitive neuroscience has relied on cross-sectional samples based on a single observation of each participant to draw connections between brain properties and cognitive skills such as reading. A brief review of the literature reveals that there are at least 27 studies on the neuroanatomical underpinnings of reading skill and of these, only 3 studies consist of more than 2 observations per participant. Furthermore, of the five large-scale, publicly-funded, developmental datasets that we are aware of, only 1 includes observations at at least four different time points. Unfortunately, a dynamic system cannot be studied with cross-sectional data - at least four time points are necessary to resolve the dynamic interactions that we discovered in the current work. The present findings provide strong evidence for the dynamic nature of brain-behavior associations and suggest that differences in white matter properties based on a single observation do not necessarily relate to stable behavioral differences. Developmental cognitive neuroscience should prioritize longitudinal, within-subjects designs that investigate brain-behavior relationships coupled with information about an individual's educational environment. This will lead to a better understanding of how different learning opportunities influence brain development and open the door for refined, developmentally appropriate learning interventions and pedagogical strategies.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The outputs of the pyAFQ tractometry pipeline, including tractography and tract profiles, are provided in a BIDS derivative directory in the FCP-INDI AWS S3 bucket. Specifically, the FA and MD tract profiles for each participants are available on S3 at s3://fcp-indi/data/Projects/HBN/BIDS_curated/derivatives/afq/combined_tract_profiles.csv. The AtlasTrack derivatives used for the ABCD and PING datasets are available through the NIH Data Archive. Instructions on how to access the PING data through the NIMH Data Archive (NDA) are provided (https://www.nitrc.org/plugins/mwiki/index.php/ping:MainPage). Instructions for accessing the ABCD data can be found here: https://nda.nih.gov/abcd/. Instructions on how to access the HCP data are provided here: https://www.humanconnectome.org/storage/app/media/documentation/LS2.0/LS_Release_2.0_Access_Instructions_June2022.pdf.

Acknowledgements

This manuscript was prepared using a limited access dataset obtained from the Child Mind Institute Biobank, Healthy Brain Network. This manuscript reflects the views of the authors and does not necessarily reflect the opinions or views of the Child Mind Institute. Additional data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. Further data used for this study were obtained from the Pediatric Imaging, Neurocognition and Genetics Study (PING) repository (https://www.nitrc.org/projects/ping), a publicly shared data resource comprising standardized assessments of behavioral, neuroimaging and genetic variables in typically developing children, adolescents and young adults (Jernigan et al., 2016). As such, the investigators within PING (T. L. Jernigan, A. M. Dale, L. Chang, N. Akshoomoff, C. McCabe, E. Newman, T. M. Ernst, P. Van Zijl, J. M. Kuperman, S. S. Murray, C. S. Bloss, N. J. Schork, W. Thompson, H. Bartsch, D. G. Amaral, E. R. Sowell, W. E. Kaufmann, P. Van Zijl, S. Mostofsky, B. J. Casey, B. Rosen, T. Kenet, J. A. Frazier, D. N. Kennedy, and J. R. Gruen) contributed to the design and implementation of PING and/or provided data but did not participate in analysis or writing of this report. Instructions on how to access the data through the NIMH Data Archive (NDA) are provided (https://www.nitrc.org/plugins/mwiki/index.php/ping:MainPage). Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development[SM] (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9–10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from https://doi.org/10.15154/1523041. This work was funded by Eunice Kennedy Shriver National Institute of Child Health and Human Development grants R01HD095861 and P50HD052120 to JDY, RF1MH121868 to AR and JDY and R01EB027585-02 to AR and Eleftherios Garyfallidis. ARH, AR and MN were also additionally supported by the Gordon & Betty Moore Foundation and the Alfred P. Sloan Foundation, through grants to the University of Washington eScience Institute Data Science Environment. Some compute resources were provided through Amazon Web Services and a grant from Microsoft Azure through the University of Washington eScience Institute.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.dcn.2024.101341.

References

Alexander, L.M., et al., 2017. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci. Data 4, 170181.
Barres, B.A., Raff, M.C.C., 1993. Proliferation of oligodendrocyte precursor cells depends on electrical activity in axons. Nature 361, 258–260.
Beaulieu, C., et al., 2005. Imaging brain connectivity in children with diverse reading ability. Neuroimage 25, 1266–1271.
Ben-Shachar, M., Dougherty, R.F., Wandell, B.A., 2007. White matter pathways in reading. Curr. Opin. Neurobiol. 17, 258–270.
Boets, B., et al., 2013. Intact but less accessible phonetic representations in adults with dyslexia. Science 342, 1251–1254.

Bringmann, L.F., et al., 2013. A Network Approach to Psychopathology: New Insights into Clinical Longitudinal Data. PLoS One 8, e60188.
Brown, T.T., et al., 2012. Neuroanatomical assessment of biological maturity. Curr. Biol. 22, 1693–1698.
Casey, B.J., et al., 2018. The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. Dev. Cogn. Neurosci. 32, 43–54.
Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16. Association for Computing Machinery, pp. 785–794.
Chiang, M.-C., et al., 2011. Genetics of white matter development: a DTI study of 705 twins and their siblings aged 12 to 29. Neuroimage 54, 2308–2317.
Chin Fatt, C.R., Jha, M.K., Minhajuddin, A., Mayes, T., Trivedi, M.H., 2021. Sex-specific differences in the association between body mass index and brain aging in young adults: Findings from the human connectome project. Psychoneuroendocrinology 124, 105059.
M. Cieslak, et al., QSIPrep: An integrative platform for preprocessing and reconstructing diffusion MRI. bioRxiv, 2020.09.04.282269 (2020).
Cieslak, M., et al., 2021. QSIPrep: an integrative platform for preprocessing and reconstructing diffusion MRI data. Nat. Methods 18, 775–778.
Deutsch, G.K., et al., 2005. Children's reading performance is correlated with white matter structure measured by diffusion tensor imaging. Cortex 41, 354–363.
Eluvathingal, T.J., Hasan, K.M., Kramer, L., Fletcher, J.M., Ewing-Cobbs, L., 2007. Quantitative diffusion tensor tractography of association and projection fibers in normally developing children and adolescents. Cereb. Cortex 17, 2760–2768.
Epskamp, S., Waldorp, L.J., Mõttus, R., Borsboom, D., 2018a. The Gaussian graphical model in cross-sectional and time-series data. Multivar. Behav. Res. 53, 453–480.
Epskamp, S., Borsboom, D., Fried, E.I., 2018b. Estimating psychological networks and their accuracy: a tutorial paper. Behav. Res 50, 195–212.
Fields, R.D., 2015. A new mechanism of nervous system plasticity: activity-dependent myelination. Nat. Rev. Neurosci. 16, 756–767.
Fieremans, E., Jensen, J.H., Helpern, J.A., 2011. White matter characterization with diffusional kurtosis imaging. Neuroimage 58, 177–188.
Fortin, J.-P., et al., 2017. Harmonization of multi-site diffusion tensor imaging data. Neuroimage 161, 149–170.
Fortin, J.-P., et al., 2018. Harmonization of cortical thickness measurements across scanners and sites. Neuroimage 167, 104–120.
Gabrieli, J.D.E., 2009. Dyslexia: a new synergy between education and cognitive neuroscience. Science 325, 280–283.
Gabrieli, J.D.E., Ghosh, S.S., Whitfield-Gabrieli, S., 2015. Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. Neuron 85, 11–26.
Garyfallidis, E., et al., 2014. Dipy, a library for the analysis of diffusion MRI data. Front. Neuroinform. 8, 8.
Germanò, E., Gagliano, A., Curatolo, P., 2010. Comorbidity of ADHD and Dyslexia. Dev. Neuropsychol. 35, 475–493.
Gershon, R.C., et al., 2013a. NIH toolbox for assessment of neurological and behavioral function. Neurology 80, S2–S6.
Gershon, R.C., et al., 2013b. NIH toolbox cognition battery (CB): measuring language (vocabulary comprehension and reading decoding). Monogr. Soc. Res. Child Dev. 78, 49–69.
Geschwind, N., 1965. Disconnexion syndromes in animals and man. I. Brain 88, 237–294.
Gibson, E.M., et al., 2014. Neuronal activity promotes oligodendrogenesis and adaptive myelination in the mammalian brain. Science 344, 480–481.
Grills-Taquechel, A.E., Fletcher, J.M., Vaughn, S.R., Stuebing, K.K., 2012. Anxiety and reading difficulties in early elementary school: evidence for unidirectional- or bi-directional relations? Child Psychiatry Hum. Dev. 43, 35–47.
Gullick, M.M., Demir-Lira, Ö.E., Booth, J.R., 2016. Reading skill-fractional anisotropy relationships in visuospatial tracts diverge depending on socioeconomic status. Dev. Sci. 19, 673–685.
M. Gumus, M.S. Kiran, Crude oil price forecasting using XGBoost in 2017 International Conference on Computer Science and Engineering (UBMK), (2017), pp. 1100–1103.
Hagler, D.J., et al., 2009. Automated white-matter tractography using a probabilistic diffusion tensor atlas: application to temporal lobe epilepsy. Hum. Brain Mapp. 30, 1535–1547.
Hagler Jr, D.J., et al., 2019. Image processing and analysis methods for the adolescent brain cognitive development study. Neuroimage 202, 116091.
Harms, M.P., et al., 2018. Extending the human connectome project across ages: imaging protocols for the lifespan development and aging projects. Neuroimage 183, 972–984.
Hasan, K.M., et al., 2012. Diffusion tensor quantification and cognitive correlates of the macrostructure and microstructure of the corpus callosum in typically developing and dyslexic children. NMR Biomed. 25, 1263–1270.
Henriques, R.N., et al., 2021. Diffusional kurtosis imaging in the diffusion imaging in python project. Front. Hum. Neurosci. 15, 390.
Hodge, M.R., et al., 2016. ConnectomeDB–sharing human brain connectivity data. Neuroimage 124, 1102–1107.
Huber, E., Donnelly, P.M., Rokem, A., Yeatman, J.D., 2018. Rapid and widespread white matter plasticity during an intensive reading intervention. Nat. Commun. 9, 2260.
Huber, E., Mezer, A., Yeatman, J.D., 2021. Neurobiological underpinnings of rapid white matter plasticity during intensive reading instruction. Neuroimage 243, 118453.
Ishibashi, T., et al., 2006. Astrocytes promote myelination in response to electrical impulses. Neuron 49, 823–832.
Jensen, J.H., Helpern, J.A., Ramani, A., Lu, H., Kaczynski, K., 2005. Diffusional kurtosis imaging: the quantification of non-gaussian water diffusion by means of magnetic resonance imaging. Magn. Reson. Med. 53, 1432–1440.
Jernigan, T.L., et al., 2016. The pediatric imaging, neurocognition, and genetics (PING) data repository. Neuroimage 124, 1149–1154.
Johnson, W.E., Li, C., Rabinovic, A., 2007. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 8 (1), 118–127.
Keller, T.A., Just, M.A., 2009. Altering cortical connectivity: remediation-induced changes in the white matter of poor readers. Neuron 64, 624–631.
Klingberg, T., et al., 2000. Microstructure of temporo-parietal white matter as a basis for reading ability: evidence from diffusion tensor magnetic resonance imaging. Neuron 25, 493–500.
Kruper, J., Yeatman, J.D., Richie-Halford, A., Bloom, D., Grotheer, M., Caffarra, S., Kiar, G., Karipidis, I.I., Roy, E., Chandio, B.Q., Garyfallidis, E., Rokem, A., 2021. Evaluating the Reliability of Human Brain White Matter Tractometry. Aperture Neuro 1 (1). https://doi.org/10.52294/e6198273-b8e3-4b63-babb-6e6b0da10669.
Langer, N., et al., 2017. White matter alterations in infants at risk for developmental dyslexia. Cereb. Cortex 27, 1027–1036.
Lebel, C., Walker, L., Leemans, A., Phillips, L., Beaulieu, C., 2008. Microstructural maturation of the human brain from childhood to adulthood. Neuroimage 40, 1044–1055.
Lebel, C., Treit, S., Beaulieu, C., 2019. A review of diffusion MRI of typical white matter development from early childhood to young adulthood. NMR Biomed. 32, e3778.
Li, J., et al., 2022. Cross-ethnicity/race generalization failure of behavioral prediction from resting-state functional connectivity. Sci. Adv. 8, eabj1812.
G. Louppe, M. Kumar, Bayesian optimization with skopt — scikit-optimize 0.8.1 documentation (2016) (December 20, 2021).
Mackey, A.P., Whitaker, K.J., Bunge, S.A., 2012. Experience-dependent plasticity in white matter microstructure: reasoning training alters structural connectivity. Front. Neuroanat. 6, 32.
Marcus, D.S., et al., 2013. Human connectome project informatics: quality control, database services, and data visualization. Neuroimage 80, 202–219.
Marek, S., et al., 2022. Reproducible brain-wide association studies require thousands of individuals. Nature. https://doi.org/10.1038/s41586-022-04492-9.
McCrimmon, A.W., Climie, E.A., 2011. Test Review: D. Wechsler Wechsler Individual Achievement Test—Third Edition. San Antonio, TX: NCS Pearson, 2009. Can. J. Sch. Psychol. 26, 148–156.
Meisler, S.L., Gabrieli, J.D.E., 2022. A large-scale investigation of white matter microstructural associations with reading ability. NeuroImage 249, 118909. https://doi.org/10.1016/j.neuroimage.2022.118909.
Moreau, D., Wilson, A.J., McKay, N.S., Nihill, K., Waldie, K.E., 2018. No evidence for systematic white matter correlates of dyslexia and dyscalculia. NeuroImage: Clin. 18, 356–366.
Niogi, S.N., McCandliss, B.D., 2006a. Left lateralized white matter microstructure accounts for individual differences in reading ability and disability. Neuropsychologia 44, 2178–2188.
Niogi, S.N., McCandliss, B.D., 2006b. Left lateralized white matter microstructure accounts for individual differences in reading ability and disability. Neuropsychologia 44, 2178–2188.
Ogunleye, A., Wang, Q.-G., 2020. XGBoost model for chronic kidney disease diagnosis. IEEE/ACM Trans. Comput. Biol. Bioinform 17, 2131–2140.
Ostrolenk, A., Forgeot d'Arc, B., Jelenic, P., Samson, F., Mottron, L., 2017. Hyperlexia: systematic review, neurocognitive modelling, and outcome. Neurosci. Biobehav. Rev. 79, 134–149.
Ozernov-Palchik, O., et al., 2019. The relationship between socioeconomic status and white matter microstructure in pre-reading children: a longitudinal investigation. Hum. Brain Mapp. 40, 741–754.
Ozernov-Palchik, O., Gaab, N., 2016. Tackling the "dyslexia paradox": Reading brain and behavior for early markers of developmental dyslexiax. Wiley Interdiscip. Rev. Cogn. Sci. 7, 156–176.
Parsa, A.B., Movahedi, A., Taghipour, H., Derrible, S., Mohammadian, A.K., 2020. Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis. Accid. Anal. Prev. 136, 105405.
Pease-Raissi, S.E., Chan, J.R., 2021. Building a (w)rapport between neurons and oligodendroglia: reciprocal interactions underlying adaptive myelination. Neuron 109, 1258–1273.
Peters, L., Bulthé, J., Daniels, N., Op de Beeck, H., De Smedt, B., 2018. Dyscalculia and dyslexia: different behavioral, yet similar brain activity profiles during arithmetic. NeuroImage: Clin. 18, 663–674.
Rasero, J., Sentis, A.I., Yeh, F.-C., Verstynen, T., 2021. Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. PLoS Comput. Biol. 17, e1008347.
A. Richie-Halford, *et al.*, An open, analysis-ready, and quality controlled resource for pediatric brain white-matter research. *bioRxiv*, 2022.02.24.481303 (2022a).
Richie-Halford, A., et al., 2022b. An analysis-ready and quality controlled resource for pediatric brain white-matter research. Sci. Data 9, 616.
Rosseel, Y., 2012. lavaan: an R package for structural equation modeling. J. Stat. Softw. 48, 1–36.
Sagi, Y., et al., 2012. Learning in the fast lane: new insights into neuroplasticity. Neuron 73, 1195–1203.
Saygin, Z.M., et al., 2013. Tracking the roots of reading ability: white matter volume and integrity correlate with phonological awareness in prereading and early-reading kindergarten children. J. Neurosci. 33, 13251–13258.
Shmueli, G., 2010. To explain or to predict? Stat. Sci. 25.
Smith, R.E., Tournier, J.-D., Calamante, F., Connelly, A., 2012. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. Neuroimage 62, 1924–1938.

Steinbrink, C., et al., 2008. The contribution of white and gray matter differences to developmental dyslexia: Insights from DTI and VBM at 3.0T. Neuropsychologia 46, 3170–3178.
Torgesen, J.K., Wagner, R., Rashotte, C., 2011. TOWRE 2: Test of Word Reading Efficiency. Pearson Clinical Assessment.
Tournier, J.-D., et al., 2019. MRtrix3: a fast, flexible and open software framework for medical image processing and visualisation. Neuroimage 202, 116137.
Tournier, J.-D., Calamante, F., Connelly, A., 2007. Robust determination of the fibre orientation distribution in diffusion MRI: non-negativity constrained super-resolved spherical deconvolution. Neuroimage 35, 1459–1472.
Turesky, T.K., et al., 2022. Home language and literacy environment and its relationship to socioeconomic status and white matter structure in infancy. Brain Struct. Funct. 227, 2633–2645.
Van Essen, D.C., et al., 2012. The human connectome project: a data acquisition perspective. Neuroimage 62, 2222–2231.
Vandermosten, M., et al., 2015. A DTI tractography study in pre-readers at risk for dyslexia. Dev. Cogn. Neurosci. 14, 8–15.
Vandermosten, M., Boets, B., Wouters, J., Ghesquière, P., 2012. A qualitative and quantitative review of diffusion tensor imaging studies in reading and dyslexia. Neurosci. Biobehav. Rev. 36, 1532–1552.
Wandell, B.A., Yeatman, J.D., 2013. Biological development of reading circuits. Curr. Opin. Neurobiol. 23, 261–268. -4.
Wang, Y., et al., 2017. Development of tract-specific white matter pathways during early reading development in at-risk children and typical controls. Cereb. Cortex 27, 2469–2485.
Wasserthal, J., Neher, P., Maier-Hein, K.H., 2018. TractSeg - fast and accurate white matter tract segmentation. Neuroimage 183, 239–253.
C. Wernicke, "Der aphasischer Symptomenkomplex: eine psychologische Studie auf anatomischer Basis. " in Wernicke's Works on Aphasia: A Sourcebook and Review., (Mouton, 1874), pp. 91–145.
Wierenga, L.M., et al., 2018. A multisample study of longitudinal changes in brain network architecture in 4–13-year-old children. Hum. Brain Mapp. 39, 157–170.
Yeatman, J.D., et al., 2011. Anatomical properties of the arcuate fasciculus predict phonological and reading skills in children. J. Cogn. Neurosci. 23, 3304–3317.
Yeatman, J.D., Dougherty, R.F., Ben-Shachar, M., Wandell, B.A., 2012a. Development of white matter and reading skills. Proc. Natl. Acad. Sci. U. S. A. 109, E3045–E3053.
Yeatman, J.D., Dougherty, R.F., Myall, N.J., Wandell, B.A., Feldman, H.M., 2012b. Tract profiles of white matter properties: automating fiber-tract quantification. PLoS One 7, e49790.
Yeatman, J.D., Wandell, B.A., Mezer, A., 2014. Lifespan maturation and degeneration of human brain white matter. Nat. Commun. 5, 4932.
Yip, S.W., Konova, A.B., 2021. Densely sampled neuroimaging for maximizing clinical insight in psychiatric and addiction disorders. Neuropsychopharmacology. https://doi.org/10.1038/s41386-021-01124-0.